© 2017. Published by The Company of Biologists Ltd | Journal of Experimental Biology (2017) 220, 2878-2886 doi:10.1242/jeb.160192

RESEARCH ARTICLE

Determining the behavioural dose–response relationship of marine mammals to air gun noise and source proximity

Rebecca A. Dunlop[1,*], Michael J. Noad[1], Robert D. McCauley[2], Lindsay Scott-Hayward[3], Eric Kniest[4], Robert Slade[5], David Paton[5] and Douglas H. Cato[1,6]

ABSTRACT

The effect of various anthropogenic sources of noise (e.g. sonar, seismic surveys) on the behaviour of marine mammals is sometimes quantified as a dose–response relationship, where the probability of an animal behaviourally 'responding' (e.g. avoiding the source) increases with 'dose' (or received level of noise). To do this, however, requires a definition of a 'significant' response (avoidance), which can be difficult to quantify. There is also the potential that the animal 'avoids' not only the source of noise but also the vessel operating the source, complicating the relationship. The proximity of the source is an important variable to consider in the response, yet difficult to account for given that received level and proximity are highly correlated. This study used the behavioural response of humpback whales to noise from two different air gun arrays (20 and 140 cubic inch air gun array) to determine whether a dose–response relationship existed. To do this, a measure of avoidance of the source was developed, and the magnitude (rather than probability) of this response was tested against dose. The proximity to the source, and the vessel itself, was included within the one-analysis model. Humpback whales were more likely to avoid the air gun arrays (but not the controls) within 3 km of the source at levels over 140 re. 1 $\mu Pa^2\ s^{-1}$, meaning that both the proximity and the received level were important factors and the relationship between dose (received level) and response is not a simple one.

KEY WORDS: Anthropogenic noise, Behavioural response study, Received level, Humpback whale, Seismic air gun

INTRODUCTION

A dose–response relationship quantifies the magnitude, or probability, of the response of an animal in relation to the dose of some stimulus or stressor. In the case of whales exposed to noise from human activities, the response may be behavioural, and the stimulus would be the noise exposure. This relationship can then, in theory, be used to predict potential impacts of an anthropogenic sound source and it is often assumed that the behavioural response will increase in magnitude as the received sound exposure level increases (e.g. Antunes et al., 2014; Goldbogen et al., 2013; Houser et al., 2013; Kvadsheim et al., 2011, 2012, 2014; Miller et al., 2012, 2014; Tyack et al., 2011; Sivle et al., 2015; Southall et al., 2012; Williams et al., 2014). Behavioural response studies are often designed with this dose–response outcome in mind. The response threshold is the minimum 'dose' required to elicit a change in behaviour and the magnitude of the response, or probability of a response, should then increase in proportion to the increase in sound exposure level, resulting in a sigmoid curve. Studies of noise-induced temporary threshold shift (TTS) in hearing sensitivity of marine mammals in captivity have produced relationships between the magnitude of TTS and the received noise energy level for several species (Finneran, 2015), typical of dose–response relationships. These have led to criteria for managing the effects of high levels of noise on marine mammals (Southall et al., 2007) as a way of avoiding hearing damage. A TTS, however, requires relatively high noise levels and thus occurs at shorter distances compared with behavioural effects, which are likely to occur at much lower levels.

So far, establishing a simple dose–response relationship between a behavioural response and noise exposure levels in marine mammals has proved elusive. In reality, this relationship is an over-simplification because of the complexity of the behavioural responses, resulting from social and environmental effects (Dunlop et al., 2013, 2015, 2016a; Ellison et al., 2012). In addition, many studies use the probability of animals responding as the response variable, rather than the magnitude of the response, and this therefore requires some sort of threshold level that separates a non-significant from a significant change in behaviour related to 'avoidance' of the source. In a dose–response study, this change in behaviour is usually related to avoidance of the source. Qualitative scoring has been used in some studies (e.g. Miller et al., 2014), where experts compared behavioural patterns during the experiment (a movement response was assumed to be an avoidance response to the noise source). A number of other studies (including the Miller et al., 2014 study) have used a more quantitative technique, called a 'change-point' analysis (https://synergy.st-andrews.ac.uk/mocha/), to determine whether each subject significantly changed its movement pattern in response to naval sonar (e.g. Antunes et al., 2014; DeRuiter et al., 2013; Miller et al., 2014). In these analyses, changes in the speed and position of whales were measured between two successive time windows and the point along the time series at which there was an obvious change in movement or dive behaviour (compared with a 'baseline' dataset) was taken to be a significant response. Further work by Antunes et al. (2014) using simulations determined the threshold at which this change in movement behaviour could be considered an avoidance response, and Curé et al. (2012) developed a 'reaction score' based on whether the animal approached or avoided the source. Other studies have categorised behavioural responses into a number of 'severity scores' (e.g. Williams et al., 2014) based on a methodology outlined in

[1]School of Veterinary Science, The University of Queensland, Gatton, QLD 4343, Australia. [2]Curtin University, GPO Box U 1987, Perth, WA 6845, Australia. [3]Centre for Research into Ecological and Environmental Modelling, University of St Andrews, St Andrews, Fife KY16 9LZ, UK. [4]School of Engineering, University of Newcastle, Callaghan, NSW 2308, Australia. [5]Blue Planet Marine, PO Box 919, Canberra, ACT 2614, Australia. [6]School of Geosciences, University of Sydney and Defence Science and Technology Group, Sydney, NSW 2006, Australia.

*Author for correspondence (r.dunlop@uq.edu.au)

R.A.D., 0000-0002-0427-6317

Received 28 March 2017; Accepted 22 May 2017

Southall et al. (2007). These scores were then used to decide 'cut-offs' between what could be considered a 'response' and what could be considered 'no response'. In all previously mentioned studies, it was then assumed that an increase in received level (dose) increased the probability that an animal would respond, thus producing a dose–response function.

To add further complexity, a dose–response relationship may not be directly due to the received level of the noise source – it may also depend on the proximity of the source to the animal. Whales may be more likely to respond to a close-by source compared with one further away, as found for beaked whales in response to different naval sonar sources (De Ruiter et al., 2013). However, the received level of a noise source is usually highly correlated with distance from the source; therefore, it is difficult to separate the effects of received level versus the effects of proximity within one experiment. To do this, different received levels are required at the same proximity, which would require different sources at different source levels.

Finally, the behavioural change may not be exclusively in response to the noise stimulus. It is possible that the whales are responding to the presence of the vessel itself, and the noise it generates (Williams et al., 2014), as well as to the noise stimulus (as found in Dunlop et al., 2015, 2016a). If undertaking experiments to test for an animal's response to a specific noise source, such as naval sonar or seismic air guns (which involve the source being towed by a source vessel), a set of controls should be undertaken to separate the effect of the experimental noise source from the effects of the vessel. Although the response to controls can be tested using a block analysis, it is more difficult to integrate any potential response to the controls in a dose–response relationship in terms of a simple regression relationship between the response and the dose. In some studies (e.g. Antunes et al., 2014; Miller et al., 2012; Sivle et al., 2015), no, or little, response to the controls was found; therefore, it was probably deemed not necessary.

In this paper, we used the previously measured response of groups of humpback whales to noise from air guns to investigate whether there is a quantifiable dose–response relationship between the received noise level and the source proximity (measured as distance from the source) and the magnitude of the behavioural response, whilst accounting for any potential effect of the source vessel itself. The study presents a novel, and comprehensive, analytical framework with which to test for a dose–response relationship which can then be used to make management decisions regarding the behavioural responses of marine fauna to anthropogenic noise sources.

MATERIALS AND METHODS

To determine the behavioural response of migrating humpback whales, *Megaptera novaeangliae* Borowski 1781, to various air gun arrays, experiments were carried out over 5 years (2010–2015). The animal ethics permit for the study was granted by the University of Queensland Animal Ethics Committee. The behavioural results of experiments one (2010) and two (2011) have been published (Dunlop et al., 2015, 2016a), and this study builds on these results by developing a dose–response model. A pre-experimental power analysis (Dunlop et al., 2012) was carried out using the behavioural response to other stimuli (Dunlop et al., 2013) as a basis for the effect size (minimum required was n=12 based on an assumption of a greater effect size than that observed in Dunlop et al., 2013).

The general experimental design has been presented before (Dunlop et al., 2015, 2016a) and will be summarised here. Experiments were performed off Peregian Beach, in Southeast Queensland, Australia (26°29′S, 153°06′E), during the migration of humpback whales from winter breeding grounds to summer feeding grounds in the Antarctic (2010 and 2011). Each experiment consisted of a series of 'trials', where a source vessel towed air guns along a designated transect in the study area while behavioural observations were made. Each trial used one of several treatments and one or two trials were conducted each good-weather day. Treatments were either 'active' where the air gun(s) was operated, or 'controls', where the air guns were towed but not operated. There were also baseline studies in the absence of the vessel.

All trials consisted of three main phases – 'before', 'during' and 'after'. In the 'before' phase, groups of whales were 'focally followed' by either land- or boat-based teams for at least 1 h. The source vessel would wait at its designated transect start point, moving the minimum necessary to keep the array astern of the vessel and to stay in the vicinity against wind and tide. In 2010, the vessel was *FV Ash Dar S*, a 19 m West Coaster, and in 2011, it was *RV Whale Song*, a 28 m, 185 ton ship.

After 1 h of the 'before' period, the 'during' period was initiated, regardless of where the focal groups were relative to the source vessel. The vessel moved along a pre-determined path (eastwards across the migration) for 1 h at 4 knots with the air gun(s) either firing or not firing while behavioural observations continued, and then stopped. Note the air gun compressor was operating during control trials as well as active trials. In 2010, a single Bolt 600B air gun with a 20 cubic inch chamber bolted into a towfish was used, towed by *FV Ash Dar S* at a depth of 5.6 m. In 2011, *RV Whale Song* towed an array of six air guns including the same 20 cubic inch gun as used in 2010 as well as three 40 cubic inch GI air guns, which enabled it to be used either as a 20 cubic inch source or as a 140 cubic inch source. There were also two 150 cubic inch air guns used in ramp-up (Dunlop et al., 2016a) but not in the analysis presented here. The spacing of the air guns and timing of firing of each element were designed to maximise the horizontal transmission of sound and reduce directional differences in radiated sound level. In both years, the air guns were towed 18 m astern of the vessels, fired every 11 s when operating and were tracked using GPS. Active trials were designated AE20 for the 20 cubic inch air gun and AE140 for the 140 cubic inch air gun, while controls were CE1 with *FV Ash Dar S* and CE2 with *RV Whale Song*.

At the end of 'during' phase, the vessel slowed again to a minimum for the 'after' period, where behavioural observations continued for another hour. The land- and boat-based observers (see below for details) were blind to both the start of the 'during' phase and whether the treatment was an active or control. The 'trial director' (the person who maintains overall coordination of the experiment, particularly the movement of the research vessels and source vessel with regard to the initiation of the trial) used a random block design approach to select which trial was to be carried out on each day (to remove any subjective choice). This allowed for a balanced sample size in terms of the number of trials carried out. On days where two trials were completed, one active and one control trial were carried out (there were never two active trials in one day).

Land-based behavioural observations were collected daily (07:00 h to 17:00 h, weather permitting) from two different stations: the northern station (32 m elevation, 100 m from the waterline) and a second station 11 km to the south (73 m-high hill called 'Emu Mountain' set 700 m back from the beach). The stations had unobstructed and overlapping fields of view. Two 'focal follow' teams (continuously observing a group for the duration of the experiment; Kavanagh et al., 2016a) were located at each station, with a third 'scan sampling' team also at the southern

Journal of Experimental Biology

station to provide contextual data (*ad lib* observations of all other groups in the area). Each team tracked humpback whale groups using a theodolite (Leica TM 1100) connected to a notebook computer running VADAR tracking software (developed by E.K.; http://www.brahss.org.au/content/vadar.html). Theodolite-derived 'fixes' were annotated with observed behaviours (e.g. blow, breach, pectoral flipper slap, tail slap, splitting apart of a group, joining together of two groups), group size and composition, and direction of travel. Groups, unless joining together, were usually separated by more than 2 km and followed a predictable course, speed and dive interval. The track lines were monitored in real-time to ensure the follow was of a single focal group. If joining, the field of view allowed these joining animals to be tracked as they approached the focal group. If splitting apart, the focal follow continued on the original group (always a female–calf). The location of singing whales in the area was determined using a fixed array of five hydrophones. Each hydrophone was attached to a mooring and connected to a surface buoy, which transmitted acoustic data to a base station ashore using a VHF transmitter. This allowed for recording and real-time tracking of vocalising (usually singing) whales at the base station (Noad et al., 2004) using differences in the arrival times of the song sounds at the different hydrophones (*Ishmael* software; http://bioacoustics.us/ishmael.html).

Three boat-based platforms (a 6 m rigid hulled inflatable and two, 5.6 and 6 m, aluminium, centre-console boats) were used for data collection of focal groups. All behaviours were recorded for each individual in the group, along with group size and group composition. The boat was tracked continuously with on-board GPS to allow the positioning of the group relative to the boat. Boats attempted to stay within 200 m of their focal group in order to maintain visibility of behaviours whilst minimising disturbance. Individual whales were identified early in the focal follow by distinctive dorsal shapes, markings and fluke markings. Groups were photo-identified throughout the focal follow and group members were constantly validated using these photos. If groups were lost, they were excluded from the analysis. The boat-based tracking of groups that were tracked simultaneously by a land-based station was validated, as land and boat data agreed. While focal groups observed from the boats were usually different to those observed from land, some were observed by both, allowing comparison of the data collected for calibrating the land-based observations (Godwin et al., 2016) and for testing effects of the small vessels on the behaviour of the groups (Williamson et al., 2016).

Received levels of air gun array and vessel noise

To calculate the received levels of air gun signals and vessel noise at the focal groups, four autonomous acoustic recording systems (CMST-DSTO noise 'loggers'; www.cmst.curtin.edu.au\products) were deployed at various positions on the sea floor throughout the area using subsea moorings with acoustic release units. The positions of the loggers were changed every few days throughout field seasons to record sounds of the air guns, vessels and ambient noise at various positions and propagation paths. Recordings were made at 27 positions, with a north–south spread of more than 20 km and an east–west spread of approximately 7.5 km. Distances of receivers from the air guns ranged from 100 m to more than 10 km. Loggers used system gains of −3 to 40 dB, with low gains used to avoid saturation of short-range air gun signals. Most systems were set with low and high gain channels, each sampling at 4 kHz. All loggers were individually calibrated by injecting and recording white noise of known level with the hydrophone in series. This allowed the full system gain with frequency response to be measured (1 Hz to Nyquist frequency). Hydrophones (either Massa TR1025C or HiTech, HTIU90) had individual calibration specifications. Eight Aquatech 520T temperature loggers were deployed (two per noise logger), one on the seabed and one 11 m above the seabed (water depths were mainly between 20 and 40 m). These were used to track water temperature across the study period to check the usual well-mixed water column at the study site as shown in CTD profiles.

Measured air gun signals recorded on the loggers were calibrated in the time domain using the full system gain, and 16 descriptive parameters defined in McCauley et al. (2003) were calculated for each signal. The air gun arrival time and source–receiver range was used to establish the source firing time and so source–receiver geometry. The received levels from the air gun arrays were measured as sound exposure level (SEL), the time integral of the squared pressure over the duration of the pulse, as defined in McCauley et al. (2003) and Dunlop et al. (2015). Source levels of the 20 and 140 cubic inch arrays were estimated as 199 and 212 dB re. 1 μPa^2 s (SEL) at 1 m, respectively.

A frequency-dependent, empirical acoustic propagation model for the study area was developed using received air gun levels recorded on the loggers (details in Dunlop et al., 2015). Water depths in the study area were mainly between 20 and 40 m so propagation over the recorded frequency range was dominated by the acoustics of the sea floor. Propagation was complicated by the presence of high-loss patches where rock was exposed on the sea floor in some areas, resulting in higher propagation loss compared with that of the surrounding seabed of relatively deep sand. These rock patches were spatially delineated using sidescan sonar (Humminbird; http://humminbird.com.au/technology/side-imaging/) and bathymetry. The propagation model was determined from regression of received levels of air gun signals as a function of range for the sand-covered areas and for the patches. The received position and arrival time of every air gun signal at followed whales were calculated from interpolation between the positions and times of consecutive whale observation points, after allowing for signal travel time. This gave the propagation path between the source and the whale for each air gun signal, and the received level was determined from the source level and the propagation loss with comparison to levels received at the nearest logger.

The noise of the source vessel during control runs was determined from the logger recordings. Alignment of the vessel GPS and logger positions allowed curves of the received noise levels of the vessels to be developed as a function of range and bearing of the receiver from the vessel. There was significant variation in the broadband received levels with bearing and this was incorporated into the estimated source levels for the vessels. Vessel noise received by whales was then determined from the source level and the propagation loss with comparison to levels received at the nearest logger.

The higher gain recordings on the loggers were used to estimate ambient noise levels for each whale observation time, i.e. the general background noise not including the sound of the source vessel or the air guns. Both ambient noise and vessel noise were measured for frequencies above 7 Hz.

Relevant results of previous studies

The following summarises results of two previous studies (Dunlop et al., 2015, 2016a,b) which form the basis of this study.

Behavioural response variables (group dive time, course deviation from south, speed of southward movement, rates of breaching, pectoral slapping and fluke slapping behaviours, and blow rate) were quantified using land- (Kavanagh et al., 2016a) and boat-based observations of focal groups. Groups that were over high-loss patches for the entire 'during' period (and therefore received relatively low levels which were difficult to quantify and which, in some cases, may not have been audible) were also excluded.

General linear mixed models (GLMMs) were fitted using R (R Development Core Team 2012) using either the *lme4* package (Bates et al., 2015) or the *glmmADMB* package (version 0.3: http://glmmadmb.r-forge.r-project.org/). The first analysis tested the response of humpback whales to a 20 cubic inch air gun (Dunlop et al., 2015) and the second to a four-stage ramp-up procedure and a constant 140 cubic inch air gun array (Dunlop et al., 2016a). Groups were found to significantly decrease their dive time and speed of southward movement and increase their course deviation from south in response to these air gun sources, with some response in the controls (Dunlop et al., 2015, 2016a). There was no evidence in either study of significant changes in rates of group surface behaviours.

The most consistent change in behaviour in the two studies was the change in movement behaviour (a decrease in the speed of southwards movement and/or an increase in course deviation from south). Therefore, for this paper, the response variable used was a measure of change in movement behaviour.

Movement measures were collapsed into two metrics. The distance between the predicted and observed positions of the group was measured after successive 10 min time bins (taking the previous 10 min of data to generate the predicted position), resulting in six measurements per group per trial (Fig. 1; difference in distance of

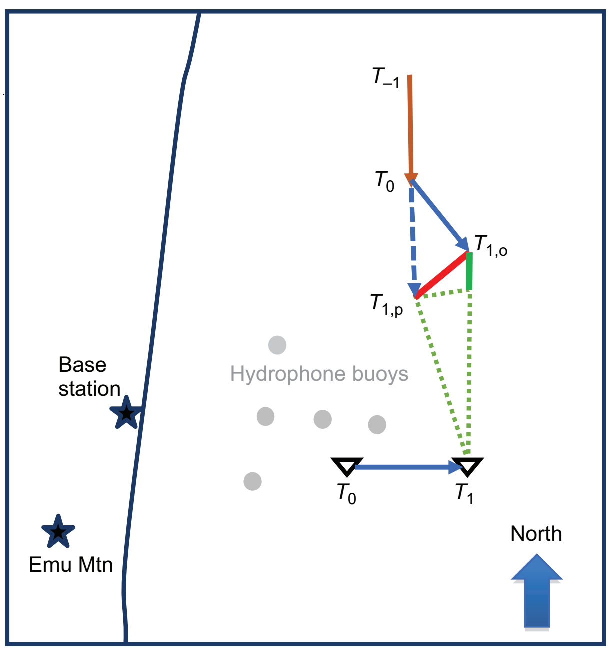

Fig. 1. Calculation of the difference in distance from the source vessel after each 10 min time bin. The position of the whale group at $T_{1,p}$ (10 min after the start of the 'during' phase at T_0) is predicted from the group's positions at T_{-1} (10 min before T_0) and T_0 (assuming no change in course and speed) and compared with where the whales were observed at $T_{1,o}$. The distance between predicted and observed positions (solid red line) is the movement deviance of the group (DD_{gp}). The distance between the group's predicted and observed position is also measured to the source vessel at T_1 (inverted triangle which started eastward at T_0). This difference in the distance to the source vessel (solid green line) is the difference in distance after 10 min (DD_{sv}).

the group, DD_{gp}). The change in predicted distance from the source vessel (Fig. 1; difference in distance to the source vessel, DD_{sv}) was also calculated after each 10 min time bin, with increased distance indicating potential avoidance and decreased distance indicating potential attraction. Maximum received levels within every 10 min were noted for both the air guns and the vessels (background noise was used if vessel noise was below ambient background noise), as well as the minimum distance of the source vessel.

To determine at what point these measures were significantly different from baseline behaviour and could be considered to be 'avoidance' or 'attraction' reactions, both the DD_{gp} and DD_{sv} were compared with similar measures for baseline data in two previous studies (Dunlop et al., 2016a,b). Observations for baseline data followed the same procedure as for the active and control trials with the position of the source vessel predicted using a simulated vessel travelling with the same speed and path as per other experimental trials (n=20 groups). These studies found that baseline groups deviated by around 200 m (95% confidence interval, 164–240 m) from their predicted course (DD_{gp}), suggesting that a group deviation from their predicted path in active and control trials could be considered to be significant if greater than 240 m (i.e. outside the confidence intervals of the baseline data). A DD_{gp} deviation of greater than 240 m equated to a DD_{sv} of greater than 150 m (towards or away from the source vessel).

The effect of treatment on these response variables was then tested using a generalised estimating equation (GEE; Hardin and Hilbe, 2002) with treatment as a factor variable (five levels: baseline, AE20, AE140, CE1, CE2). In response to the 20 cubic inch source, the 140 cubic inch source and one of the controls (CE2), groups deviated significantly further from their predicted course compared with baseline groups (DD_{gp}>240 m). However, there was no significant difference found between the DD_{sv} in baseline groups compared with the control and active groups. This is because DD_{sv} responses ranged from negative (indicating the group approached the source by >150 m) to positive (indicating the group avoided the source by >150 m). Therefore, despite significant changes in movement behaviour (DD_{gp}), at a population level groups did not consistently approach or avoid the source, resulting in a DD_{sv} close to zero for each treatment. The dose–response analysis presented here therefore uses received level and proximity as continuous predictor variables within each treatment, rather than using treatment as a factor variable.

Testing for a dose–response relationship

This dose–response analysis aimed to determine whether the response magnitude was related to received level and/or proximity to each source (20 cubic inch air gun, 140 cubic inch array and two different control treatments using two different source vessels. A generalised additive model framework was used to model the response variables DD_{gp} and DD_{sv} using R software (R Core Development Team 2015). Specifically, the MRSea (Scott-Hayward et al., 2014) and geepack (Yan and Fine, 2004; Højsgaard et al., 2006) packages were used for model fitting and selection.

DD_{gp} and DD_{sv} were calculated for all control groups (n=20) and active groups (n=31) as per previous studies. These groups were north of the source vessel start point at the start of the 'during' phase, and moving in a southward direction. Groups that were south of the vessel at the start of the 'during' phase were excluded as were those that were 'milling' (high course and speed deviation but remaining in one area) as the behaviour of groups that are already past the source and that of non-migrating (socialising) groups is likely to be

Journal of Experimental Biology (2017) 220, 2878-2886 doi:10.1242/jeb.160192

different. The response data ranged from 0 to 1800 (DD_{gp}) and from −959 to 1135 m (DD_{sv}) and a Gaussian distribution was appropriate. Environmental covariates wind speed and water depth were analysed (found to be significant variables in predicting movement behaviour by Kavanagh et al., 2016a,b, in a study on the normal migratory behaviour of humpback whales at this site) along with treatment [factor variable with four levels: AE20 (n=15 groups), AE140 (n=16 groups), CE1 (n=8 groups), CE2 (n=12 groups)], received level (SEL) and source vessel proximity (SVP, measured as distance to the source vessel). Wind and depth were considered as smooth terms and an interaction term was assessed between SEL and SVP.

To fit the interaction between SEL and SVP, both continuous covariates, a complex region spatial smoother (CReSS; Scott-Hayward et al., 2014) was used to fit a 2D smooth surface, with a spatially adaptive local smoothing algorithm (SALSA; Walker et al., 2011). The surface was modelled as the sum of basis functions that decayed exponentially with distance from a series of points termed 'knots'. The rate of exponential decay of the basis functions was chosen to show local patterns without overfitting. CReSS modelling incorporates both the active treatments with air guns operating and control treatments with the vessel towing the air guns while they are silent (the main source of noise for the controls being the vessel). The exponentially decaying nature of the basis functions ensures that response predictions are localised in terms of received level and proximity; in other words, the different decay functions ensure that control and active treatments are spatially separated. As the received noise levels of the vessel are much less than those of the air guns for the same proximity, they are well separated on the surface, so that the modelled dose–response is local to each source and determined only by that particular source. Hence, the dose–response prediction for an air gun noise source is independent of that for a vessel source.

Smooth terms were fitted using degree 2 B-splines. Owing to the differing units for the two interacting covariates, they were both scaled and centred for analysis. For both the 1D and 2D smooths, Bayesian information criteria were used for selection of number and location of knots. The 2D smooth (the relationship between proximity and received level for the different noise sources) was then used as one of the covariates within the analysis. Others included were water depth and wind speed, given that both variables were found to be significant predictors of movement behaviour in previous studies (Dunlop et al., 2015, 2016a; Kavanagh et al., 2016b), and treatment. Model selection of covariates was undertaken using fivefold cross-validation, where a smaller number indicated a better fitting model. Once the optimal model was selected, it was rerun in a GEE (Hardin, 2005) framework to deal with the lack of independence of model residuals. Focal ID was chosen as the panel structure, within which residuals are permitted to be correlated, and between which they are considered independent. The independent working correlation structure was used to calculate robust standard errors, which in the presence of positive correlation are inflated compared with the raw standard errors. These robust standard errors do not affect parameter estimates but allow appropriate calculation of confidence intervals. Predictions were made from the best model and a parametric bootstrap from the GEE model was used to calculate 95% confidence intervals and presented in figures.

RESULTS

The final model for the DD_{gp} included a linear effect of wind speed, water depth and the interaction term, and that for the DD_{sv} included a linear effect of wind and the interaction term (in R notation):

$$DD_{gp} \sim s(\text{SEL, SVP, d.f.} = 4) + \text{wind speed} + \text{water depth}, \quad (1)$$

$$DD_{gp} \sim s(\text{SEL, SVP, d.f.} = 4) + \text{wind speed}. \quad (2)$$

Treatment was not selected in either response model, probably because of the large variance in the reaction to each treatment. The interaction between proximity and the received level of the various noise sources (CReSS term) was a significant predictor in the group movement deviation (Fig. 2) and the deviation of groups away from the source vessel (Fig. 3). Spatially, within this term, the only region where animals avoided this source was where the noise was from air guns (upper left of the graph), received levels were over 140 dB re. 1 μPa^2 s or dB re. 1 μP (where 1 μP=1×10^{-7} Pa s) and the source was less than 3 km away. Response data (including 95% confidence intervals) are displayed in Figs 2 and 3.

An avoidance response at this distance was not evident in the control trials, suggesting any avoidance was in response to the air guns rather than to the source vessel itself.

These results can then be used to create the usual dose–response plots, whilst controlling for the effect of proximity (Fig. 4A) or received level (Fig. 4B). For example, at 2 km from the vessel (Fig. 4A), the DD_{sv} response to the air gun noise was predicted to be 100 m (95% confidence intervals of 40–150 m) and modal received SELs were 149 and 160 dB re. 1 μPa^2 s for the 20 cubic inch air gun and 140 cubic inch air gun array, respectively. The spread of data (from 130 to 160 dB re. 1 μPa^2 s) was due to the different propagation pathways of the air gun signals. There was no obvious difference in response between the two air gun sources at this distance despite the 11 dB difference in modal received level at 2 km. The difference in response between air guns is more to do with the combined effects of received level and proximity. Fig. 4B illustrates the effect of proximity for the same received level produced by the two different sources. Despite received levels being the same, groups responded more to the 20 cubic inch source, which would have been about 2 km closer to the groups at this received level.

DISCUSSION

Southerly migrating groups of humpback whales (which were moving towards the air gun source) have been shown to respond to various air gun sources by slowing their progression southwards (Dunlop et al., 2015) and deviating more from their southwards course (Dunlop et al., 2016a). Given that these response variables were all related to the movement of the group, they were collapsed into one measure related to their distance from the source vessel, resulting in a measure of displacement behaviour (Dunlop et al., 2016a,b). This displacement measure was significantly greater in control and active groups compared with baseline groups (which were measured in relation to a simulated vessel), suggesting general avoidance of the source vessel (though in some cases groups approached the source vessel). Using this one measure of movement behaviour allowed the magnitude of displacement to be modelled against received level and source proximity with the inclusion of the controls. By using two air gun sources with significantly different source levels, the correlation between received level and proximity was broken, allowing their effects to be separated.

CReSS, a spatial smoother model (Scott-Hayward et al., 2014) was used to fit a surface of the response variable in terms of the received level and proximity (Figs 2 and 3). The smoothness of the surface was chosen to show local effects without overfitting. As the

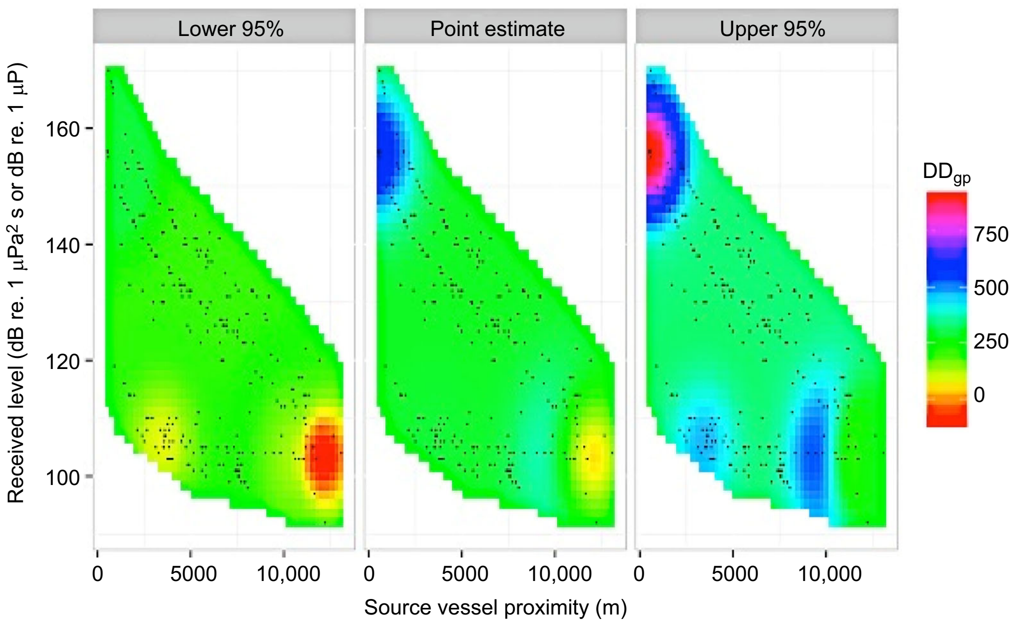

Fig. 2. Estimated DD_{gp} according to the received level of different sources of noise at a range of proximities. The colour represents the estimated response, whilst the dots give the measured received level and proximity values (controls are in the lower band and the two air gun sources are in the two upper bands). The complex region spatial smoother (CReSS) term accounts for the different relationships between the received levels of the four different sources of noise [140 cubic inch air gun array (*n*=16 groups), 20 cubic inch air gun (*n*=15 groups), CE1 control with *Ash Dar S* (*n*=8 groups) and CE2 control with *Whale Song* (*n*=12 groups)] and proximity through its local radial basis functions and provides separate estimates of response for air guns compared with controls.

received levels from the air guns are much higher than those from the vessels for the same proximity, they are well separated, and the localised nature of the surface fitting ensures that the response of each source is not affected by the response of other sources. In other words, the different noise sources, with different relationships between level and proximity, were used to create the surface. Then the response to each separate noise source was evaluated in the model output, which used the surface in place of the original measures. Significant responses to the air guns occurred when the source was within about 3 km and the received level was greater than about 140 re. 1 µPa^2 s. When controlling for received level, groups responded more to the smaller source (which was closer) than to the larger source, illustrating that proximity to the source is also important (as suggested in De Ruiter et al., 2013). It should be noted that these values do not represent the threshold of response but that responses were more likely to occur within these bounds than outside them. In addition, the response was highly variable in that some groups did not respond within these values while others responded outside them. Treatment (when used as a fixed effect in a block analysis-type framework) was not a significant effect for either response model (DD_{gp} or DD_{sv}) because of this large variation in response. Some groups avoided the source, whilst others approached within the same

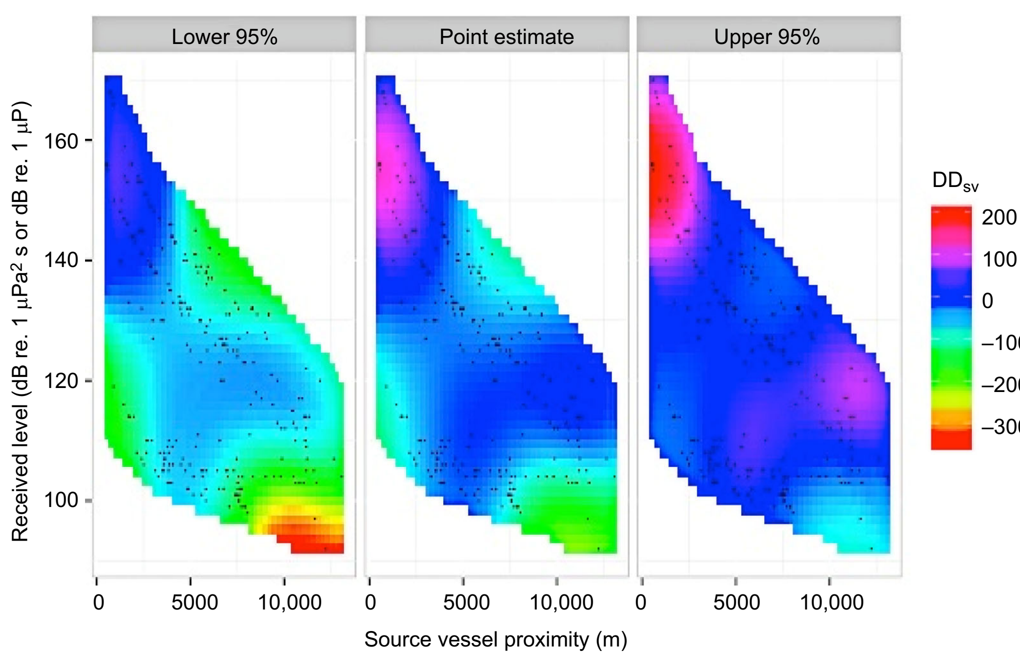

Fig. 3. Estimated DD_{sv} according to the received level of different sources of noise at a range of proximities. The colour represents the estimated response, whilst the dots give the measured received level and proximity values (controls are in the lower band and the two air gun sources are in the two upper bands). The CReSS term accounts for the different relationships between the received level of the four different sources of noise [140 cubic inch air gun array (*n*=16 groups), 20 cubic inch air gun (*n*=15 groups), CE1 control (*n*=8 groups) and CE2 control (*n*=12 groups)] and proximity through its local radial basis functions and provides separate estimates of response for air guns compared with controls.

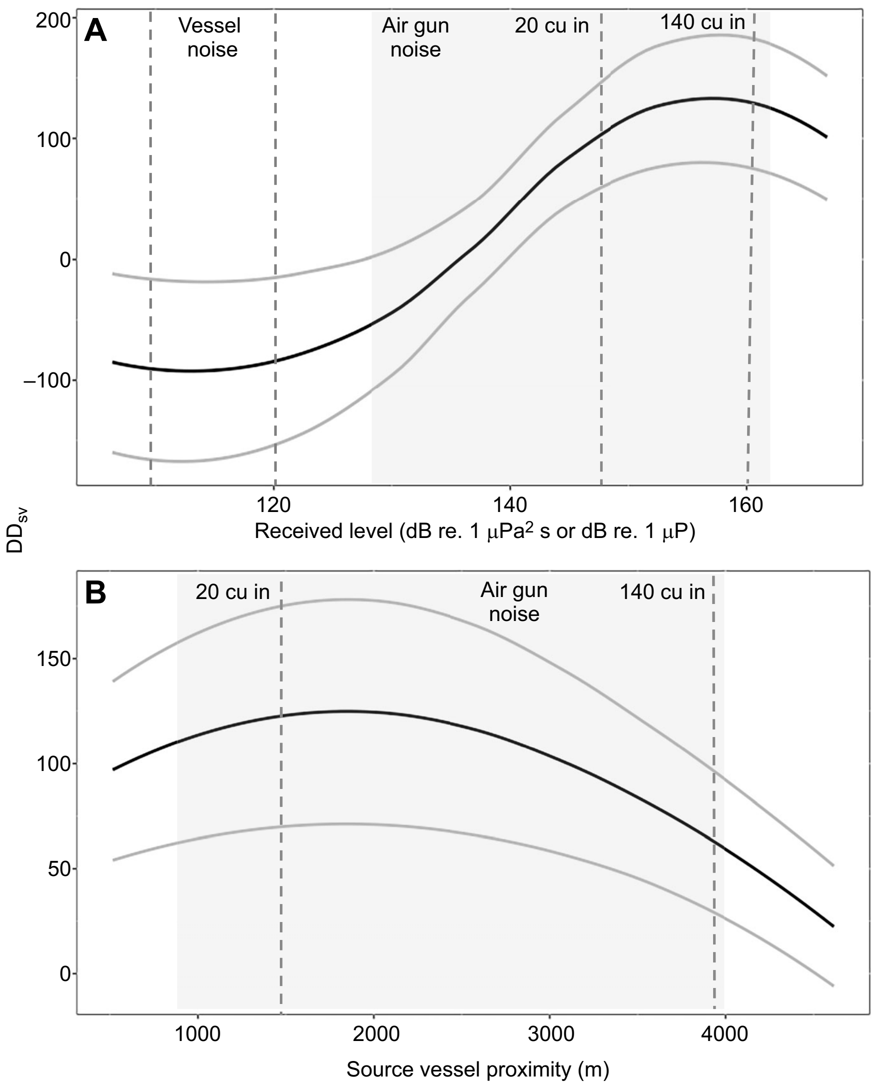

Fig. 4. Dose–response plots. DD_{sv}, including 95% confidence intervals, as a function of received level controlling for the effect of proximity (2 km from the vessel; A) and vessel proximity controlling for the effect of received level (150 dB re. 1 μPa^2 s; B). The dashed vertical lines indicate the received levels for the two air gun sources at 2 km (A; *n*=31 groups) and the proximity of the two sources at a received level of 145 dB re. 1 μPa^2 s (B; *n*=31 groups), and the grey indicates the data limits for the air gun noise (with the vessel noise noted in A).

treatment, resulting in the population DD_{sv} for each treatment being around zero. In other words, not all movement responses translated into an avoidance response; therefore, a change in movement behaviour should not be assumed to be avoidance of the source.

Despite this variance, groups did exhibit a typical dose–response in that an increase in received level resulted in a stronger avoidance response. However, it should be noted that the avoidance response within the active treatments was relatively small, in that groups deviated from their predicted course by an estimated 500 m, resulting in about a 100 m difference in distance from the source vessel. Typical humpback whale groups at this study site travel at an average speed of 4 km h^{-1} (Noad and Cato, 2007). If, however, the group responded dramatically by swimming north away from the source at 12 km h^{-1} (within the range of speeds measured by Noad and Cato, 2007), the maximum deviation of the group from its predicted pathway would be approximately 2670 m (over 10 min), showing that the whales are capable of a much greater deviation than observed. It is therefore important to put any response into context. In this study, even the outlier groups displayed a deviance from their original course of only 1500 and 1800 m, and these two groups approached the source vessel, causing a 'shut-down' (where the experiment was halted because groups reached SELs of over 170 dB re. 1 μPa^2 s per shot). Apart from these two groups, the maximum group deviation from their predicted pathway was just over 1000 m, resulting in a 750 m greater distance from the source vessel.

No clear dose–response to the controls (the vessels) was found. Although a behavioural response to the controls was found in previous studies (and it was therefore deemed necessary to include controls in the dose–response analysis), this was highly variable in terms of its occurrence, and the proximity and received levels of the groups that did respond. In addition, Dunlop et al. (2016a) suggested any response to the controls was short term and occurred within the first 20 min. Therefore, the response to the controls did not translate into a consistent response that could be related to received level and/or proximity to the ship. The temporal inconsistency of these responses precluded the development of a meaningful dose–response relationship. Other studies investigating changes in behaviour of marine mammals in response to increased levels of naval sonar also considered the effect of treatment (control versus active) in their initial analysis, though not in the final dose–response analysis (Antunes et al., 2014; Miller et al., 2012; Sivle et al., 2015). Little response to the controls was found in these studies, though the differences in study design (such as the trajectory of the source relative to the groups, response variables measured, behavioural state of the whales, species, vessel size, etc.) make it difficult to compare results.

Journal of Experimental Biology (2017) 220, 2878-2886 doi:10.1242/jeb.160192

Our study design, where the source vessel followed a predetermined path, meant that it was independent of the positions of the whale groups and they were not at a uniform distance and bearing to the source vessel at the beginning of the 'during' phase. This is in contrast to several other whale behavioural response studies that target individual whales by heading towards them up to a specified range or increase the level of a stationary source (Antunes et al., 2014; De Ruiter et al., 2013; Goldbogen et al., 2013; Miller et al., 2014; Southall et al., 2012; Tyack et al., 2011). These are known as 'dose–escalation' studies. The disadvantage of our approach is that it allows less control of exposure levels so there are fewer data points at higher received levels. The advantage, however, is that by not controlling this, the movement of the source is independent of the movements of the whales, so is more realistic, as this is typical of actual seismic surveys and sonar operations. In addition, this study is only applicable to migrating whales approaching a source vessel that is moving directly across their migratory path, although the whales do show significant behaviour typical of the breeding grounds. Despite this, the results of this study are surprisingly consistent with previous studies with humpback whales in different behavioural contexts. Feeding humpback whales, for example, responded at ranges up to 3 km from the source, at levels of 150–169 dB re. 1 µPa (Malme et al., 1985). Resting female humpback whales with calves displayed avoidance reactions at 140 dB re. 1 µPa, though other cohorts reacted at higher levels (157–164 dB re. 1 µPa; McCauley et al., 2003).

This study presents a methodology with which to assess any dose–response relationship between a noise source and a behavioural response in marine mammals. It includes any potential effects of the vessel itself (by including controls where the air guns were towed by the vessel but not operated) and tests both the effect of proximity to the source and received level. We found that both received level and proximity of a source are important in terms of eliciting a response. It should be noted that the values found here (3 km and 140 dB re. 1 μPa^2 s.) are specific to this context, sound type and behavioural state of the whales, and therefore may be different in other contexts. Also, some groups did not respond within these values, while others responded outside them, so they do not provide an absolute threshold. However, the results give dose–responses in terms of both received level and source distance in a realistic scenario that will be useful in mitigation and designing ramp-up. Both received level and proximity to the source should be considered when making management decisions regarding the mitigation of the interaction between whales and noise sources.

Acknowledgements
We would like to thank all the BRAHSS (Behavioural Response of Australia Humpback whales to Seismic Surveys) team members particularly Louise Bennett and Verity Steptoe and the BRAHSS volunteers for their assistance with data collection. We would also like to thank Ailbhe Kavanagh, Michael Williamson, Elise Godwin and Sasha Collecutt for their assistance with preliminary data analysis and Simon Blomberg for his advice on statistical analysis. We thank Curt and Michelle Jenner for providing *RV Whale Song* and Chandra Salgado Kent for leading the source vessel observer team. Finally we would like to thank the MOCHA team, based at the Centre for Research into Ecological and Environmental Modelling (CREEM), specifically Monique MacKenzie, for their valuable input regarding the statistical analysis.

Competing interests
The authors declare no competing or financial interests.

Author contributions
Conceptualization: R.A.D., M.N., R.M., D.C.; Methodology: R.A.D., M.N., R.M., E.K., R.S., D.P., D.C.; Formal analysis: R.A.D., R.M., L.S.-H.; Investigation: R.A.D.; Data curation: R.A.D., M.N., R.M., L.S.-H., E.K., R.S., D.P.; Writing - original draft: R.A.D., M.N., D.C.; Writing - review & editing: R.A.D., M.N., R.M., L.S.-H., E.K., R.S., D.P., D.C.; Project administration: R.A.D., M.N., R.M., D.C.; Funding acquisition: R.A.D., M.N., R.M., D.C.

Funding
Funding was provided as part of Joint Industry Programme on E&P Sound and Marine Life, managed by the International Association of Oil & Gas Producers (IOGP). The principal contributing companies to the programme are BG group, BHP Billiton, Chevron, ConocoPhillips, Eni, ExxonMobil, IAGC, Santos, Statoil and Woodside. The US Bureau of Ocean Energy Management (BOEM), Origin Energy, Beach Energy and AWE Limited provided support specifically for the BRAHSS study. Funding agencies had no input into experimental design, data collection, data analysis or interpretation or the writing of the paper.

References
Antunes, R., Kvadsheim, P. H., Lam, F. P. A., Tyack, P. L., Thomas, L., Wensveen, P. J. and Miller, P. J. O. (2014). High thresholds for avoidance of sonar by free-ranging long-finned pilot whales (*Globicephala melas*). *Mar. Pollut. Bull.* **83**, 165-180.

Bates, D., Mächler, M., Bolker, B. M. and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1-48.

Curé, C., Antunes, R., Samarra, F., Alves, A. C., Visser, F., Kvadsheim, P. H. and Miller, P. J. O. (2012). Pilot whales attracted to killer whale sounds: acoustically-mediated interspecific interactions in cetaceans. *PLoS ONE* **7**, e52201.

DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone, E. A., Friedlaender, A. S., Joseph, J. E., Moretti, D., Schorr, G. S. et al. (2013). First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. *Biol. Lett.* **9**, 20130223.

Dunlop, R. A., Cato, D. H. and Noad, M. J. (2012). Behavioral response studies: problems with statistical power. In *The Effects of Noise on Aquatic Life*, Vol. 730 (ed. A. N. Popper and A. Hawkins). Series title: Advances in Experimental Medicine and Biology, pp. 293-297. New York: Springer Science+Business Media.

Dunlop, R. A., Noad, M. J., Cato, D. H., Kniest, E., Miller, P. J. O., Smith, J. N. and Stokes, M. D. (2013). Multivariate analysis of behavioural response experiments in humpback whales (*Megaptera novaeangliae*). *J. Exp. Biol.* **216**, 759-770.

Dunlop, R. A., Noad, M. J., McCauley, R. D., Kniest, E., Paton, D. and Cato, D. H. (2015). The behavioural response of humpback whales (*Megaptera novaeangliae*) to a 20 cubic inch air gun. *Aquat. Mamm.* **41**, 412-433.

Dunlop, R. A., Noad, M. J., McCauley, R. D., Kniest, E., Slade, R., Paton, D. and Cato, D. H. (2016a). Response of humpback whales (Megaptera novaeangliae) to ramp-up of a small experimental air gun array. *Mar. Pollut. Bull.* **103**, 72-83.

Dunlop, R. A., Noad, M. J. and Cato, D. H. (2016b). A spatially explicit model of the movement of humpback whales relative to a source. In Effects of Noise on Aquatic Life, Dublin, 10–16 July 2016. Proceedings of Meetings on Acoustics (POMA), Acoustical Society of America, 010026, 13 pp.

Ellison, W. T., Southall, B. L., Clark, C. W. and Frankel, A. S. (2012). A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds. *Conserv. Biol.* **26**, 21-28.

Finneran, J. J. (2015). Noise-induced hearing loss in marine mammals: a review of temporary threshold shift studies from 1996 to 2015. *J. Acoust. Soc. Am.* **138**, 1702-1726.

Godwin, E. M., Noad, M. J., Kniest, E. and Dunlop, R. A. (2016). Comparing multiple sampling platforms for measuring the behavior of humpback whales (*Megaptera novaeangliae*). *Mar. Mamm. Sci.* **32**, 268-286.

Goldbogen, J. A., Southall, B. L., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Hazen, E. L., Falcone, E. A., Schorr, G. S., Douglas, A., Moretti, D. J. et al. (2013). Blue whales respond to simulated mid-frequency military sonar. *Proc. R. Soc. B Biol. Sci.* **280**, 20130657.

Hardin, J. W. (2005). *Generalized estimating equations (GEE)*. *Encyclopedia of Statistics in Behavioral Science*. Chichester: John Wiley & Sons.

Hardin, J. W. and Hilbe, J. M. (2002). *Generalized Estimating Equations*. Boca Raton, FL: Chapman & Hall, CRC Press.

Højsgaard, S., Halekoh, U. and Yan, J. (2006). The R Package geepack for generalized estimating equations. *J. Stat. Softw.* **15**, 1-11.

Houser, D. S., Martin, S. W. and Finneran, J. J. (2013). Exposure amplitude and repetition affect bottlenose dolphin behavioral responses to simulated mid-frequency sonar signals. *J. Exp. Mar. Biol. Ecol.* **443**, 123-133.

Kavanagh, A. S., Goldizen, A. W., Blomberg, S. P., Noad, M. J. and Dunlop, R. A. (2016a). Factors affecting the reliability and validity of behavioural datasets: assessing the impact of observers' experience and native language on studies of wild animals. *Aquat. Mamm.* **42**, 1-11.

Kavanagh, A. S., Noad, M. J., Blomberg, S. P., Goldizen, A. W., Kniest, E., Cato, D. H. and Dunlop, R. A. (2016b). Factors driving the variability in diving and movement behavior of migrating humpback whales (*Megaptera novaeangliae*): implications for anthropogenic disturbance studies. *Mar. Mamm. Sci.* **33**, 413-439.

Kvadsheim, P., Lam, F. P. and Miller, P. J. (2011). Behavioural response studies of cetaceans to naval sonar signals in Norwegian waters-3S-2011 cruise report. Published: 2011. URL: http://rapporter.ffi.o/rapporter/2011/01289.pdf.

Kvadsheim, P., Lam, F. P. and Miller, P. J. (2012). Behavioural response studies of cetaceans to naval sonar signals in Norwegian waters-3S-2012 cruise report. Published: 2012. URL: http://rapporter.ffi.no/rapporter/2012/02058.pdf.

Kvadsheim, P., Lam, F. P., Miller, P., Wensveen, P., Visser, F., Sivle, L., Oudejans, M., Kleivane, L., Curé, C., Ensor, P. et al. (2014). Behavioural responses of cetaceans to naval sonar signals-the 3S-2013 cruise report. FFI-rapport 2014/00752. Published: 2014. URL: http://rapporter.ffi.no/rapporter/2014/00752.pdf.

Malme, C. I., Miles, P. R., Tyack, P., Clark, C. W. and Bird, J. E. (1985). *Investigations of the Potential Effects of Underwater Noise from Petroleum Industry Activities on Feeding Humpback Whale Behavior*. Cambridge, MA, U.S: Report 5851 from Bolt Beranek & Newman Inc. Minerals Management Service, Anchorage, AK.

McCauley, R. D., Fewtrell, J., Duncan, A. J., Jenner, K. C. S., Jenner, M. N., Penrose, J. D., Prince, R. I. T., Adhitya, A., Murdoch, J. and McCabe, K. (2003). Marine seismic surveys: analysis and propagation of air-gun signals; and effects of exposure on humpback whales, sea turtles, fishes and squid. In *(Anon) Environmental Implications of Offshore Oil and Gas Development in Australia: Further Research*, pp. 364-521. Canberra: Australian Petroleum Production Exploration. http://cmst.curtin.edu.au/wp-content/uploads/sites/4/2016/05/McCauley-et-al-Seismic-effects-2000.pdf.

Miller, P. J. O., Kvadsheim, P. H., Lam, F. P. A., Wensveen, P. J., Antunes, R., Alves, A. C., Visser, F., Kleivane, L., Tyack, P. L. and Sivle, L. D. (2012). *The severity of behavioral changes observed during experimental exposures of killer (Orcinus orca*), long-finned pilot (*Globicephala melas*), and Sperm (*Physeter macrocephalus*) whales to naval sonar. *Aquat. Mamm.* **38**, 362-401.

Miller, P. J. O., Antunes, R. N., Wensveen, P. J., Samarra, F. I. P., Alves, A. C., Tyack, P. L., Kvadsheim, P. H., Kleivane, L., Lam, F. P. A., Ainslie, M. A. et al. (2014). Dose-response relationships for the onset of avoidance of sonar by free-ranging killer whales. *J. Acoust. Soc. Am.* **135**, 975-993.

Noad, M. J. and Cato, D. H. (2007). Swimming speeds of singing and non-singing humpback whales during migration. *Mar. Mamm. Sci.* **23**, 481-495.

Noad, M. J., Cato, D. H. and Stokes, M. D. (2004). Acoustic tracking of humpback whales: measuring interactions with the acoustic environment. Proceedings of Acoustics 2004, Annual Conference of the Australian Acoustical Society, Gold Coast, 3–5 November, pp. 353-358.

Scott Hayward, L. A. S., MacKenzie, M. L., Donovan, C. R., Walker, C. and Ashe, E. (2014). Complex region spatial smoother (CReSS). *J. Comput. Graph. Stat.* **23**, 340-360.

Sivle, L. D., Kvadsheim, P. H., Curé, C., Isojunno, S., Wensveen, P. J., Lam, F.-P. A., Visser, F., Kleivane, L., Tyack, P. L., Harris, C. M. et al. (2015). Severity of expert-identified behavioural responses of humpback whale, minke whale, and northern bottlenose whale to naval sonar. *Aquat. Mamm.* **41**, 469-502.

Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Green, C. R., Jr, Kastak, D., Ketten, D. R., Miller, J. H. and Nachtigall, P. E. (2007). Marine mammal noise exposure criteria: initial scientific recommendations. *Aquat. Mamm.* **33**, 411-521.

Southall, B. L., Moretti, D., Abraham, B., Calambokidis, J., DeRuiter, S. L. and Tyack, P. L. (2012). Marine mammal behavioral response studies in southern california: advances in technology and experimental methods. *Mar. Technol. Soc. J.* **46**, 48-59.

Tyack, P. L., Zimmer, W. M. X., Moretti, D., Southall, B. L., Claridge, D. E., Durban, J. W., Clark, C. W., D'Amico, A., DiMarzio, N., Jarvis, S. et al. (2011). Beaked whales respond to simulated and actual navy sonar. *PLoS ONE* **6**, e17009.

Walker, C. G., Mackenzie, M. L., Donovan, C. R. and O'Sullivan, M. J. (2011). SALSA – a spatially adaptive local smoothing algorithm. *J. Stat. Comput. Simul.* **81**, 179-191.

Williams, R., Erbe, C., Ashe, E., Beerman, A. and Smith, J. (2014). Severity of killer whale behavioral responses to ship noise: a dose-response study. *Mar. Pollut. Bull.* **79**, 254-260.

Williamson, M. J., Kavanagh, A. S., Noad, M. J., Kniest, E. and Dunlop, R. A. (2016). The effect of close approaches for tagging activities by small research vessels on the behavior of humpback whales (*Megaptera novaeangliae*). *Mar. Mamm. Sci.* **32**, 1234-1253.

Yan, J. and Fine, J. P. (2004). Estimating equations for association structures. *Stat. Med.* **23**, 859-880.